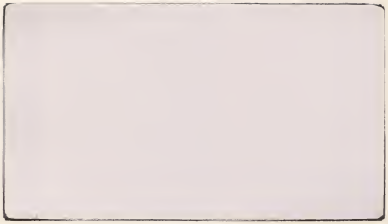

MANDEX, INC.

DEVELOPING MD-DRG ALGORITHMS

Howard West, Project Director
Peter McMenamin, Consultant
Leo Marcus, Statistician

Federal Project Officer: William Sobaski

February 6, 1985

HCFA Contract No. HCFA-500-84-0024

MANDEX, INC.
8302D Old Courthouse Road
Vienna, Virginia 22180

TABLE OF CONTENTS

Initial Aims and Project Development

The project was originally conceived as a feasibility study involving the computer matching of HCFA in-house Part A and Part B records for hospitalized beneficiaries. A one percent sample would be constructed involving Part A records from the MEDPAR data files based on all beneficiaries with HIC numbers with the two-digit ending sequence 95. These HIC numbers would be matched against the Bill Summary Record file of Part B claims abstracts to prepare a merged file.

The particular focus of the study was the exploration and development of algorithms to assess by DRG the patterns of physician services which accompanied hospitalizations. A specific problem that was envisioned involved the set of services that could be considered as part of a hospital episode but that were provided on an ambulatory basis just prior to or just following hospitalization. Second, assuming that appropriate "windows" of service had been identified with respect to beneficiaries who had had only a single hospitalization, algorithms would be explored to adapt these windows to the analysis of the claims experience of beneficiaries with two or more hospitalizations in the same year.

Coincidentally, Mandex had been contracted by ASPE to conduct a feasibility study of the potential for merger of carrier and intermediary data from two states (South Carolina and Florida). The focus of that effort was more directly on the question of estimation of the amounts of physician services which could be associated with specific DRGs.

The two projects proceeded somewhat symbiotically with results from one giving added perspective to the examination of the other. And, over the life of the two projects, both were directed to become more focused on the production of policy relevant statistics with respect to physician reimbursement system reforms that might be linked to the DRG-based hospital prospective payment reimbursement system being implemented for Part A.

Project Reports

All of the statistics and recommendations which have come from the study have been reported to HCFA, most notably in the report prepared for the ASPE contract titled, Physician and Hospital Reimbursement Study. In addition, Mandex submitted written interim progress reports and provided a variety of oral presentations of interim findings. For the purpose of summarizing the project in this report, we will provide a commentary on the interim reports to highlight some of the lessons to be learned from working with these data. In addition, the Executive Summary from the Physician and Hospital Reimbursement Study and the texts of the interim progress reports have been included as appendices.

Commentary

-- Early in the project it was determined that one percent sample files would be prepared for the states of South Carolina and Florida to allow beneficiary specific comparison of claims data from both the carrier/intermediary files and HCFA's in-house files. Items missing from one file but not the other often revealed information about the development of each of the files that could not have been identified looking at one file only. A remarkable finding was the predominant correspondence between the two sets of data, suggesting that the HCFA in-house data was substantially complete.

-- The most puzzling finding, and one which significantly changed the development of the project, involved the lack of "hospitalization related" ambulatory care services for substantial proportions of beneificiaries across all DRGs. Initially, it was believed that this might have been associated with DRG miscodings of surgical admissions into medical DRG categories. Investigating this possibility identified many of the ways in which miscodings may be -- and were -- made based on the Part A data only, but it did not resolve the problem of the apparently missing services. This remains a question for future research. The potential magnitude of this currently unacknowledged Medicare liability could be substantial, and any "cost saving" physician

reimbursement reform which happened to facilitate submission of bills for those "missing" services might prove to be very expensive.

-- Review of both Part A and Part B data for each DRG from a random sample of 24 cases from single admissions in South Carolina gave the first clues about the problems of DRG miscoding in the MEDPAR data. Surgeries which were not coded or which were miscoded on the Part A claim were clearly apparent in the CPT coding on the Part B claims. For the most part, this failure to correctly code the surgery resulted in assignment into the relevant Medical DRG based only on the diagnosis. Also observed were within-surgery miscodings caused by relatively ambiguous coding options within ICD-9-CM.

In addition to "surplus" surgeries in Medical DRGs, there were cases of omitted surgery. For the most part, these involved date or place of service errors in the claims records. However, there were some cases where no surgery bill could be found regardless of date or place of service. The initial Mandex recommendation was to delete these latter cases with no surgery claims when computing average charges. However, given the evaluation that the files were relatively complete, omission of surgical cases without surgery claims from these computations might lead to estimates of <u>average</u> Medicare physician reimbursements by DRG which exceed actual average reimbursements per case for some DRGs.

-- Given the relative plethora of data because of access to 100% of the intermediary/carrier records, adjustments for presumed data errors took the form of eliminating observations rather than trying to correct them. The choice was made for several reasons. First, examination of the data showed that such errors occurred with sufficient frequency to warrant some action, but not with sufficient frequency to invalidate the remaining data. Second, corrections for data omissions in the procedure coding of the South Carolina hospital records were fairly successfully carried out on a case-by-case basis using the Part B records, but there were still some cases where there was insufficient information in those records to make a non-arbitrary correction. Given the costliness of that effort, data filtering seemed the better option.

Others, notably Mitchell, et al., have chosen the route of data correction rather than filtering.

-- Although perceived to be a potentially difficult problem, the incidence of cases where HCFA Bill Summary Records could not be uniquely assigned to one or two or more hospitalizations was much smaller than anticipated. A 68% rule of thumb was found to give a good estimate of how many cases could be expected to have only unambiguous Part B Bill Summary Records with respect to specific hospitalizations. That is, roughly 68 percent of all beneficiaries with two stays only would have no ambiguous Bill Summary Records. For beneficiaries with three stays only, (68%) * (68%) = 46% would have no ambiguous Bill Summary Records. For four stays, $(68\%)^3$ would be the estimator, etc. This would imply that approximately half of the Part A claims of beneficiaries with multiple stays could be used without having to develop additional algorithms for assigning potentially ambiguous Bill Summary Records.

-- There are a variety of options available for assignment of any potentially ambiguous records. Two design considerations influenced the Mandex recommendations. First was the need for economy which led to a recommendation for single pass options rather than multiple pass options. That is, having to examine each ambiguous Part B record only once without regard to the disposition of any other ambiguous or unambiguous Part B records might save a significant amount of data processing without introducing significant additional error. Second, the decision to not introduce unnecessary error led to a recommendation that any algorithms which are developed involve unique assignments of ambiguous claims rather than partitions or proportional allocations of those claims to each of the potentially matched Part A episodes. There should be no marked differences between the estimates of average Part B allowed charges that might be developed using each of these two approaches, but the proportional allocation approach clearly involves an assignment error with respect to each Part B claim rather than some fraction of those claims.

INTERIM PROGRESS REPORTS

March 14, 1984
April 12, 1984
June 15, 1984
July 15, 1984
July 23, 1984
August 22, 1984
October 22, 1984
November 22, 1984
January 22, 1985

March 14, 1984

Progress During the First Month

A meeting at the call of the Project Officer was held in Baltimore immediately following award of the contract. All of the HCFA staff that have been involved with development of the project and in matching and analyzing the 1981 1% sample of the MEDPAR and Bill Summary Record Files participated in the discussions. A variety of problems were explored. These had come to light during the process of matching the Medicare in-patient records (Part A) to the Part B bill records that make up the Bill Summary Record data system. A number of tables were reviewed that were derived from the matching of the bill records for persons having a single episode of hospital care during the year.

A primary objective of the meeting was to explore the most effective and least intrusive modus operandi for the key HCFA staff and the Mandex staff to accommodate their parallel and over-lapping efforts. It was agreed that all requests for data and/or tabulations by Mandex staff would be made to the Project Officer who would assume a coordinating role. This would reduce to the maximum extent any duplication of effort and assure that the experiences of the HCFA group that has focused on single admissions and that of the Mandex group that will focus on multiple admissions will serve each other.

The first task facing the contractor is to ascertain the extent and character of the data matching problems, and to develop recommendations regarding them. Mandex has 1981 Medicare data for two states, South Carolina and New Jersey. The South Carolina data have been matched (Part A stay records with Part B bills) with a match rate of 98%. It was decided that the most straightforward approach in attacking the problems with the HCFA 1% sample data would be to select the MEDPAR and Bill Summary Record data for these two states from the national data file. These selected records would be compared with the same 1% sample of South Carolina intermediary data in Mandex's possession. Accordingly, Mandex requested a tape copy of the HCFA file for these two states. This was received on March 7. Mandex has selected and printed out all of the South Carolina records for HIC numbers ending in 95, the matching 1% sample. It was decided to delay selection of the New Jersey 1% sample until New Jersey Blue Cross resubmitted its Medicare Part A tape for 1981, including the missing DRG codes. This tape was received on March 12. Not all New Jersey hospitals were reimbursed on the basis of DRGs during 1981. Thus there will be some voids in the New Jersey Part A data elements since no diagnosis or procedure codes are contained in the Blue Cross stay records for that state.

Progress planned during the next month will include the matching and evaluation of the two sets of data to try to determine the extent and scope of differences. Also, the significance of any omissions and/or apparent inaccuracies will be evaluated. To the extent feasible, the algorithms to be developed will need to incorporate adjustments designed to allow for deficiences in the basic data.

April 12, 1984

Progress During the Second Month

Our efforts during the month have been focused on the comparison of the HCFA 1% Medpar, Bill Summary Record data for South Carolina with a 1% sample of the South Carolina Blue Cross and Blue Shield 1981 Medicare data drawn by Mandex using the HCFA 1% sample digits (95). A printout was prepared in HIC number order of the inpatient records for each sample person followed by all of the matching Part B bills with each set in date sequence. The tape copy of the 1% Medpar records for South Carolina was also listed out in HIC number order with multiple admissions in date sequence.

These two sets of records were now matched on a line-by-line (record-by-record) basis to determine the extent of failure to match. As part of the matching process, these records were compared as to DRG assignment and total covered charges. Total covered charges proved to be the best data element for matching of specific stays for patients having multiple stays.

An exact match of all of the hospital stay records was not possible for a number of reasons:

1. The Medpar file includes 29 stay records with admissions in 1980 and discharges in 1981. All such records were entirely excluded from the "Mandex" file.

2. Two hospitals (provider numbers 420050 and 420079) do not use South Carolina Blue Cross as their intermediary. Provider No. 420050 had 11 stays in the 1% Medpar file and provider No. 420079 had 29 stays in the 1% Medpar file.

3. The "Mandex" file lacked 47 stays, each terminating in December in the Medpar file. This is being investigated.

4. The HIC number for Railroad Retirement beneficiaries (there appear to be 16 RR records in the Medpar file and 11 RR records in the "Mandex" file), although not matchable by computer, could be matched by us by inference and some speculation. Incidentally, only three Part B bills are included for these RR beneficiaries in the "Mandex" (South Carolina Blue Shield) file.

Except for these four conditions, matches were made between all but two of the stay records in the two files.

Two observations need to be noted about the Medpar DRG assignments. First, 137 stays were assigned to DRG 470 as opposed to none in the Mandex file. Second, there are 204 other differences in DRG assignment between the two files. The assignments to DRG 470 are caused by the presence of an "x" in the fifth position in the Medpar diagnosis codes when a fifth digit is called for in the ICD-9-CM coding system. The 204 conflicts in assignments of DRGs are frequently caused (138 occurrences) by the presence of

an "x" in the fourth position of the surgical procedure code when a fourth digit is called for in the ICD-9-CM coding system. The remaining conflicts are due to a variety of factors; e.g., use of age rather than month, day, year of birth; lack of a secondary diagnosis, etc.

The 1% Bill Summary Record sample (BSR) and the "Mandex" 1% Part B sample for South Carolina were also compared. This was accomplished by matching these two sets of records to determine whether one or more BSR and Part B records were present for each HIC number contained in the matching "Mandex" stay record. This process determined that there were five HIC numbers for which no BSR was present. This is a negligible finding and contributes but little to explaining the 6 to 10% reported failure to match Medpar with BSR files. However, it is our understanding that the matching criteria include both HIC number and month of service. Therefore, the 29 1980-1981 Medpar stay records must be added to the unmatched group, since there are no 1980 BSR records in that file. Also, the 16 RR HIC numbers in the Medpar file could not match the six probable RR HIC numbers in the BSR files. Thus the total of failures to match the Medpar sample is 50 (5 missing, 29 1980-81, and 16 RR), or 4% of the 1202 Medpar records. The remainder undoubtedly are due to month-to-month non-matches. Such may well occur when an inclusive dates bill for a series of hospital visits is dated on the first month of the series and all charges are assigned to that month when the stay extends into the next month (or vice versa, depending on billing practice), and there will be no BSR bill to match that month.

Problem Avoidance

Avoidance of the problems caused by the lack of a fifth digit in the ICD-9-CM diagnosis code when a fifth digit is required by the code system structure can be accomplished by the following processes:

> Identify all ICD-9-CM diagnosis codes that require a fifth digit. The DRG grouper will reject each record having a four-digit diagnosis code when a fifth digit is called for. Assign, by computer, as the fifth digit a zero to all codes rejected by "grouper." Grouper will accept most but will reject a small proportion of these zero assignments. Eliminate these zeros and assign a one to the fifth digit of these rejected records. All but a negligible number of cases will now be accepted by the DRG grouper as having valid diagnosis codes. A zero in the fifth digit is used for the "unspecified" term in a series of terms. A one in the fifth digit is usually a specific term among a short series of related diagnostic terms.*
>
> A second approach would be to examine the tabular list, Volume 1 of ICD-9-CM, 9th Revisions, DHHS Publication No. (DHS) 80-1260.* Terms requiring a fifth digit can be identified and the appropriate fifth digit selected so as to meet the requirements of the DRG assignment system. A machine readable list of these could be prepared for use whenever the DRG grouper rejects records for lack of fifth digits.

* See pages 296 and 297 of the Tabular List for examples of the use of zero and one as fifth digits.

Attached are two lists based on our work with 1981 South Carolina Part B data that will help avoid problems caused by the lack of a fourth digit in the ICD-9-CM procedure code. List A consists of three digit procedure codes that require a fourth digit together with a fourth digit that will assign the record to the appropriate surgical DRG. List B consists of three-digit procedure codes that require a fourth digit which cannot be asigned by pre-selection. Records containing these codes should be automatically assigned by appropriate means to DRG 471 (unknown surgical procedure).

Both List A and List B lack 100% of all ICD-9-CM procedures requiring a fourth digit, based as they are on the experience in one state for a single year. However, they can serve as a point of departure for development of such complete lists.

Work Planned During the Third Month

We are planning to develop two estimating methods during the next thirty days. The first is the estimating method for the proportion and value of claims excluded by the calendar year nature of the HCFA sample files. The second is the general estimating method for refining calculation of the values of the MD-DRGS.

LIST A

ICD-9-CM Code	Assigned 4th Digit	ICD-9-CM Code	Assigned 4th Digit	ICD-9-CM Code	Assigned 4th Digit
013	1	112	2	195	2
015	2	113	2	200	1
020	2	114	2	204	1
021	1	115	1	205	1
023	1	120	0	207	1
030	2	121	1	210	4
032	1	126	4	217	2
033	2	127	2	223	1
041	9	130	0	224	1
047	1	131	1	225	0
049	2	134	1	226	0
058	9	135	9	227	1
060	2	136	1	230	1
063	1	137	0	231	1
065	0	142	1	237	0
068	1	143	1	259	4
072	1	144	9	262	1
076	1	145	4	263	0
079	1	147	2	264	1
081	1	151	1	272	1
082	0	152	1	277	1
083	8	160	1	281	1
085	1	163	1	289	1
086	1	164	1	300	1
087	0	165	1	322	1
088	2	166	1	333	4
091	1	168	1	334	2
098	1	182	1	340	2
099	1	183	9	345	1
102	1	187	1	350	0
103	1	191	1	351	0
104	1	192	1	352	0

ICD-9-CM Code	Assigned 4th Digit	ICD-9-CM Code	Assigned 4th Digit	ICD-9-CM Code	Assigned 4th Digit
353	1	464	0	551	1
356	0	465	0	552	4
357	0	466	0	553	1
359	1	468	0	555	1
361	0	482	1	558	1
377	0	483	5	559	7
382	1	484	9	563	3
401	1	486	1	564	0
404	0	488	1	565	1
413	3	489	1	566	1
419	3	490	1	567	1
420	9	494	6	571	2
421	0	495	9	574	9
422	1	499	5	575	9
423	2	501	2	577	1
424	0	510	2	579	3
425	1	511	3	589	9
428	3	513	1	592	1
429	1	517	1	597	1
434	2	519	3	599	1
438	1	520	1	601	2
439	1	521	2	609	3
440	0	525	1	614	2
441	1	530	0	619	9
443	1	531	0	621	2
444	0	532	1	624	1
451	1	533	1	635	1
452	1	534	9	638	1
455	0	535	1	641	1
456	2	536	1	644	9
457	5	538	0	649	2
460	1	549	9	651	1
461	0	550	1	652	2

ICD-9-CM Code	Assigned 4th Digit	ICD-9-CM Code	Assigned 4th Digit	ICD-9-CM Code	Assigned 4th Digit
655	1	753	6	829	1
656	1	761	1	830	2
659	1	764	3	834	9
662	1	766	1	839	1
663	1	767	0	851	2
665	1	775	1	853	1
671	1	802	6	854	1
673	1	810	0	855	3
681	4	813	1	860	9
691	1	815	1	869	1
692	3	817	1	875	3
695	1	820	9	885	2
709	1	821	1	922	7
711	1	822	1	950	4
712	4	824	5		
716	2	826	1		
717	1	828	1		

LIST B

ICD-9-CM Procedure Code Numbers

011	319	469	705	801
012	332	487	707	804
024	339	491	710	807
029	342	497	749	808
039	347	502	760	809
040	349	506	763	811
044	371	512	769	812
052	372	514	770	814
061	373	515	771	816
069	378	516	772	818
084	380	529	773	819
094	381	541	774	823
116	383	542	776	831
117	384	546	778	832
123	385	547	779	833
124	386	556	780	836
125	388	572	784	837
128	392	573	785	838
129	393	578	786	840
169	394	584	787	841
202	395	590	789	842
209	399	606	791	844
218	402	666	792	849
273	405	682	793	852
274	446	690	794	858
275	450	699	795	862
295	453	701	796	866
302	454	702	798	867
312	459	703	800	868
317	467			

YEAR END CROSSOVERS

Year end crossovers, i.e., beneficiaries admitted to a hospital in one year but discharged in the next, will be a problem for estimation because of the calendar year nature of the files. The Scope of Work calls for an identification of the proportions and physician service values of stays which were crossovers.

Proportions

There is little reason to believe the the proportion of stays which are crossovers changes from year to year. Crossovers as a percentage of January discharges -- alternately, as a percentage of December admissions -- should be comparable in 1980, 1981, 1982. etc. Estimating these proportions would simply require enumerating both the crossovers and all January discharges (or December admissions). Fortunately, by accident, January 1981 discharges with 1980 admissions were included in the MEDPAR file received with respect to South Carolina. Hence it would be possible to estimate that percentage. All December admissions, regardless of crossover status, were deleted from the file, hence that percentage cannot be estimated directly.

Physician Service Values

Estimating the physician service values of crossover stays may be more problematic in the absence of a two-year file of data. On the one hand, for a crossover beneficiary admitted to a hospital, statistics from the average course of treatment for his DRG are probably good estimations of actual treatments. Hence the non-crossover data might be used to estimate physician service values for crossover stays. On the other hand, beneficiaries admitted to hospitals in, say, the last ten days of December are more likely to be sicker than average, otherwise their admissions might be postponed til January after the Christmas/New Years vacation periods. Hence average physician service values might be minimum estimates. (There is also a difference in case mix which could be anticipated for crossovers, since beneficiaries receiving elective surgery probably do schedule their hospitalizations to avoid that time period. Severity is more of a problem in estimating than case mix.)

Again, in the absence of a two-year data file, there is no direct way of estimating these values. One might compare average same month post discharge physician service values for January crossover discharges to all other January and February discharges to determine whether there are any significantly higher values for the January crossovers.

PHYSICIAN SERVICE VALUES

The notion behind the study of "physician DRGs" is that there may be relatively homogeneous packages of physician services which accompany the hospital services provided during a hospitalization for a specific DRG. In fact, the model behind the origin of the work at hand was one in which a patient saw his or her physician on an ambulatory basis. During one of these visits, one or more diagnostic tests would be performed. Given certain results from those tests, hospitalization might be recommended. In many of those cases, hospitalization would occur. During the hospitalization, the patient would receive additional tests and/or physician services. And following discharge, there would be one or more follow-up visits.

Reimbursing physicians for a package that included only the physician services provided in hospital would run the risk of not accounting for some services which might be performed either in hospital or on an ambulatory basis. Further, if rates were set based on existing practice, some services currently performed in hospital might be shifted to an ambulatory setting, in effect potentially creating a situation of double reimbursement. Hence the interest in examining "windows" of service immediately preceding and following a hospitalization.

Given the episode model of hospitalization, a straightforward approach would be to identify characteristic patterns of care for specific DRGs. For example, one would estimate the average (alternately, the median) number of days between the initiation of ambulatory care and the date of admission and between the date of discharge and a break in ambulatory service sufficient to delineate the end of an episode.

Two problems have been identified to pursuing this approach: a mechanical one involving the Bill Summary records date of service and an empirical one involving patterns of billing.

Service

For 1981, Bill Summary Record data indicate only the month of service. Therefore identification of windows of service with these data will be constrained. One could estimate average physician service values by DRG for the month (or two months) preceding the admission and the month or months following discharge. Allocating physician service values for non-inpatient service during the month(s) of hospitalization will be problematic. Where the admission and discharge dates are in separate months, the allocation between pre-admission and post-discharge periods is unambiguous. Where the admission and discharge are in the same months (probably 2 times out of 3), the allocation is more complex. We have identified two possible approaches: 1) simply allocate half of that month's physician service values to pre-admission, half to post-discharge; or 2) allocate the month's physician service values by the ratio of days pre-admission to days post-discharge using the MEDPAR admission and discharge dates. The differences between these two methods and the validity of each might be assessed using the original carrier data from South Carolina and/or Florida which have specific dates of service -- not merely month of service.

Zero Service Patterns

A greater conceptual problem has been identified empirically which may obviate the need to estimate windows: in a great many cases, there may be no windows. For example, based on South Carolina data on single admissions for 27 of the most frequently occurring DRGs, the minimum percentage of patients who did not see an attending physician in the 15 days prior to hospitalization was 52 per cent (where an attending physician was any physician who billed for services performed during the time frame of the hospitalization). Extending that window to 60 days and including any physician who saw the patient prior to the hospitalization, a maximum of 86 percent of patients (hospitalized for lymphoma or leukemia) were billed for physician services during that time period. However, that implies that roughly one leukemia patient in seven was not. For hospitalized cases of atherosclerosis more than one third of the patients had no physician claim within 60 days prior to the admission.

Comparable statistics can be found on the post-discharge side. For both lymphoma and leukemia (DRG 404) and pacemaker implantation (116) nearly 81 per cent of patients had a physician billing record for the period up to 60 days following the discharge. That is, one in five did not. For digestive malignancies for patients 69 years old or younger only 44 percent of patients had a physician billing record for the 60 day post-discharge period.

Some of the post-discharge surgical patterns may be explained by surgeons' global bills. Any reimbursements for follow-up visits are included in the surgical fee billed as of the date of the surgery. For example, for lens procedure, only 10.9 percent of patients had a post-discharge bill from one of their attending physicians. However, for medical treatments of respiratory neoplasms (DRG 82), only 22.9 percent of patients have any physician claim for the 60 days post-discharge period, and only 45.8 percent have any claim from any physician in that period. Global bills cannot explain those low percentages.

In the absence of a finding of homogeneity in aggregate physician service values before, during, and after hospitalization, these statistics suggest that it would be more fruitful to examine the homogeneity of within hospital physician service values before trying to ascertain appropriate windows on either side.

June 15, 1984

Progress during the Fourth Month

The attached "Review of the Data" provides a fairly detailed description of activities and the data problems dealt with during the period. Attachment 2 of the Review provides a series of seven recommended steps that apply to the HCFA Single Episode Analysis.

Work was begun on development of algorithms for apportioning values of physicians' services in cases of multiple admissions. However, we want to validate these by testing them with available data and were unable to complete this step.

As we examine more of the detailed data, we are constantly reminded that many of the "single" admissions would be multiples if a 12-month exposure period rather than a calendar year were the time frame for analysis. And, of course, coverage limits and missing BSR or MEDPAR data are also factors.

REVIEW OF THE DATA

Several potential problems have been identified through the examination of both non-matches' Health Insurance Print Outs (HIPO) and aggregate matched beneficiaries' Bill Summary Records (BSR) data by DRG. Non-matching is the first problem. The problems identified through the BSR statistics include excess surgery in medical DRGs, surgical DRGs apparently without surgery, and relatively high co-efficients of variation in surgical charges per case by DRG.

Non-matches

HIPO data were requested for 125 of the non-matched cases: beneficiaries with MEDPAR records for 1981 but apparently no BSR records. Of those, 17 had no B coverage at all in 1981; 8 began B coverage in 1981 in some month besides January, 1 terminated coverage in 1981; 12 died in 1981. The latter may only partially explain non-matches, particularly with respect to non-assigned claims. Four beneficiaries were Railroad Retirement Board beneficiaries, 3 are cross references, and 3 had indications of black lung benefits. Two were enrolled in HMOs. One had a B1 HIC number suffix, and three have no record of a hospital discharge in 1981 although 2 of those were admitted in 1981. Eleven beneficiaries apparently did not exceed the deductible. All told, just about half of the non-matches have a plausible reason for non-matching.

Two of the non-matches were South Carolina residents and could be checked against the South Carolina Blue Cross Blue Shield file. One did not become eligible for B until December 1981; her hospitalization was in March. Bills for her for March do appear in the South Carolina file with zero reimbursement; there are no records in the BSR file. The other South Carolina resident does have claims in both in SCBCBS file and in the 1% BSR file, hence is not a true non-match. Marilyn Newton indicated this may have occurred if the BSR records were out of HIC sequence.

Of the remaining half of the non-matches without plausible rationalizations for not matching, there are no obvious patterns that have been identified. The

distribution of 1981 discharge dates, however, does exhibit a puzzling bimodal distribution with a primary mode in February (and March) and a secondary mode in November (and October). It was hypothesized that discharges in the last quarter of the calendar year might have had physician claims submissions too late for inclusion in the 1981 BSR files. The primary mode in February cannot be explained that way.

Excess Surgery in Medical DRGs

The single admission data review captured all BSR data with an inpatient place of service recorded in any month which did not precede the hospital admission month. In many of the medical DRGs relatively expensive amounts of surgery were indicated in the BSR records, suggesting surgery more complex than the various endoscopies which could have been recorded as surgical services.

For each DRG, a sample of 24 beneficiaries was drawn (or all beneficiaries if fewer than 24 were hospitalized in South Carolina). Reviewing these sets of complete B records suggests two major reasons for the apparently excess surgery: second hospitalizations occurring in December which were not included in the MEDPAR file and miscoding of surgery in hospital bills leading surgical cases to be categorized into medical DRGs. (There are also cases where surgical treatments are miscoded into other surgical DRGs.) (Attachment 1 includes a summary of the reviews of several DRGs using the 24 case samples.)

Missing December claims are more of a nuisance than anything else. The upshot of the error is merely that beneficiaries with multiple admissions are categorized as singles and December (and some January) surgery may be counted against medical DRGs. Of selected medical DRGs with 24 cases: 1 is missing a December A record in DRG 82, respiratory neoplasms; 2 in DRG 172, digestive malignancy with DX 2, 1 in DRG 173 digestive malignancy without DX 2 (a January A claim is missing for another); 1 in DRG 88, chronic obstructive pulmonary disease; 1 in DRG 89, pneumonia; 2 in DRG 182, gastroenteritis, etc., with DX2; 1 in DRG 183, gastroenteritis, etc., without DX 2. This problem is common but can be corrected fairly simply.

The miscoding problem has more substantial effects for prospective payment. In all probability, average estimated hospital costs are not affected for surgical DRGs which "lose" the misclassified cases to medical categories. Estimated average costs for those medical DRGs, however, should be biased upward. Most of the cases seem to be errors of omission: there is no record of a surgery on the hospital claim. For example, under medical treatments of respiratory neoplasms, one case had a lobectomy with separate B claims from surgeon, assistant surgeon, and anesthesiologist. This should have been included in DRG 75. (DRG 75 has a weight of 2.5851 compared to DRG 82 with 1.0713.) Three cases under medical treatments of digestive malignancies with DX 2, DRG 172, have Part B claims for colectomies or colostomies suggesting they should have been classified in DRG 148. (The weights for DRGs 148 and 172 are 2.6601 and 1.278, respectively.) The 24 digestive malignancies without second diagnoses (DRG 173) include 5 that might have been classified into surgical DRGs 116, 149, or 154. (The weights for DRGs 116, 149, 154, and 173 are 2.8574, 2.2706, 2.9984, and 1.0888, respectively.)

Under the various medical treatments of cardiovascular problems (DRGs 120-140), the Part B claims reveal there are many patients who have received pacemaker implants, pulmonary surgery, amputations, skin grafts, treatments of skull fracture, and one total hip replacement. Again, typically the weight of the "proper" DRG is at least twice that of the medical DRG into which it was misclassified.

A comparable problem also occurs involving misclassifications of surgical patients into other surgical DRGs. This is especially evident in "wastebasket" types or "other" categories. For example, DRG 233 is "other muskuloskeletal and connective tissue operating room procedures with DX 2." Review of the Part B records suggests that out of 24 sample cases in this category, 1 should be in DRG 209 (with a total hip replacement), 1 in DRG 213 (amputation), 1 in DRG 216 (bone biopsy), 1 in DRG 218 (shoulder athroplasty), 1 in DRG 222 (fractured kneecap), 1 in DRG 225 (hammertoe) and 3 in DRG 229 (hand surgery). For this DRG, the net effect on the estimated average costs due to the misclassifications would be unclear. DRGs 209 and 213 have higher weights than DRG 233. The remainder have lower weights.

Missing Surgery

HCFA's printouts suggest a substantial incidence of surgical DRG discharges that do not have BSR records of inpatient surgery. It was not uncommon to find 6 to 9 percent of the discharges by DRG without a physician record of inpatient surgery. The data do not suggest that the problem is that great in the state of South Carolina, but they do suggest some of the reasons for the problem: misstated place of service and/or date of service. For a variety of DRGs, one can find surgery where the place of surgical service is listed as the physician's office even though the date of service coincides with the patient's hospitalization dates. This also occurs with place of service listed as an outpatient department. It is not clear whether this is associated with individual physicians' billing styles or is random error. The date of service misstatements probably are associated with batched bills covering several dates. In one sample of 24 discharges, one physician's claim for a pacemaker implantation had the same date of service as an office visit -- 3 months before the hospitalization.

There appear to be a nonnegligible number of cases where there is no bill for surgery properly or improperly coded as to place and date of service. In some of these cases, there is a claim from an anesthesiologist or assistant which is consistent with the DRG, in other cases there is no claim related to the presumed surgery. These might involve charity care, professional courtesy, teaching care, other insurance, out of state physicians, unfiled bills, or missing bills. However, there are no obvious associations between these omissions and other factors which might be relevant to the treatment. Hence eliminating such cases from consideration is not unwarranted.

Relatively Large Coefficients of Variation

The coefficient of variation is the ratio of the standard deviation of a distribution to its mean. As a rule of thumb, distributions with coefficients of variation greater than 1.0 are described as exhibiting relatively great variability. (Although there is no statistical reason for this -- even with normal distributions, there are an infinite number of distributions with a coefficent of variation greater than one.)

Many of the single episode DRGs in the 1% sample are observed with coefficients of variation of surgical charges per stay which exceed 1.0. This also occurs with respect to other types of service. This could imply a policy problem, for example, in implementing single rates per DRG; an analysis problem in trying to retain commensurate cases in specific DRGs; a data problem, viz., missing or incorrect data; or some combination of all three problems.

Review of the detailed Part B records of 24 cases per DRG suggests several sources of variation. These include differences between individual physicians in charges for the same procedure and differences in procedures being performed within the same DRG. For example within DRG 125, Cardiac Catherization in Circulatory Disorders without AMI, there is a three-fold variation in the submitted charges for the cardiac catherization depending on whether its site is the aorta or the coronary artery. The between procedure charge variation probably contributes more to the observed large coefficients of variation than between physician differences.

The coefficient of variation calculations can be explored mathematically to identify the conditions which would yield a coefficient of variation in excess of 1. The calculations yield a condition involving the ratio of the sum of the squares of the observations to the sum of all of the crossproducts: the greater the ratio, the larger the coefficient of variation. A relatively large numerator would occur with respect to distributions with a relatively fat tail of outliers on the high side. A relatively low denominator would occur with respect to distributions with a signficant number of zero valued observations.

This latter appears to be more likely the source of the relatively large coefficients of variation. In many categories and particularly for surgery, a significant percentage of the bill summary record tables indicate zero charges by category of service.

This is partly a data problem with respect to some missing data and/or misclassified DRGs. Observations that are questionable in this regard can be censored from the data set to resolve this problem. (A listing of recommendations

suggested to HCFA is included as Attachment 2. The recommendations have been communicated to Ira Burney.) The analysis problem remains. Dealing with variations in practice and/or billing style will be more complicated. Finally, there may or may not be a policy problem related to this variation depending on whether there is some uniformity in aggregate charges.

ATTACHMENT 1
Selected DRGs - Review of Sample Cases

DRG-121-140 - Medical treatment of cardiovascular diagnoses

DRG 121-124 - Circulatory disorders with and without AMI - Several cases appear to have pacemaker implantation CPT codes -- should be DRG 116

DRG 125 - Uncomplicated circulatory disorders without AMI with cardiac catherization - Cardiac catherization coded on A claim in lieu of several major surgical procedures not coded: bypass, valve replacement, pacemakers, skull fracture.

Two thirds (16) of the discharges come from hospital number 4, five from hospital 26, three hospitals each had one discharge. The most common submitted charge for the cardiac catheterization was $549 in hospital 4 (13 times out of 16) for a coronary artery catherization. Three different practices billed this amount.

DRG 127-128 - Heart failure and shock, and deep vein thrombophlebitis, respectively - Mainly medical as would be expected

DRG 129 - Cardiac arrest - One major pulmonary surgery

DRG 130-131 - Peripheral vascular disorders - Several endarectomies, 3 amputations

DRG 132-134 - Atherosclerosis and hypertension (134) - Several cardiac catherizations not recorded on A claim.

DRG 88-90 - Chronic obstructive pulmonary disease (88) and pneumonia - Several cases with missing A claims.

DRG 182-183 - Esophagitis, etc. - Several cases with missing A claims. Several scopies, many excisions of skin lesions.

DRG 197 - Cholecystectomy -- 100% with surgery in hospital.

DRG 290 - Major Joint -- Seven knees, seventeen hips. One knee with pacemaker. One hip with anesthesia only. Coding nonconcordence: (surgeon = total hip, anesthesia = not), three times - (anesthesia = total hip, surgeon = not), twice; (hospital = total hip, surgeon and anesthesia = not), once. Same sort of problem in DRG 211: five cases with total hip on surgeon and/or anesthesia claim but not A claim.

Attachment 1 - Continued

DRG 210-211 - Hip and femur not including major joint. Seven of 48 cases missing a surgery claim or place of service for surgery were coded as OPD.

DRG 336, 116, 161-162 - Transurethral prostatectomy (336), pacemaker implant (116) and hernia procedures - Several missing second A. Several with skin lesion excisions. Several with date of service nonconcordance with hospital stay.

DRG 233-234 - Other muskuloskeletal and connective tissue O.R. procedures. A wastebasket. Many cases where imprecise ICD-9-CM codes were used. Three missing surgery claims and/or surgery place of service coded as office.

DRG 235 - Fractured femur — Medical. Several cases could be DRG 209 or 211.

DRG 236 - Fractured hip or pelvis — Medical. Several cases might have been DRG 211

DRG 82 - Respiratory Neoplasms — Medical. Three with B claims for during period only, all with out of state HIC numbers (one died). Eight cases with no care post hospital.

DRG 172-173 - Digestive malignancies — Medical. Eight live discharges with no care post hospital in both 172 and 173. Five deaths in each category. Several that might have been DRG 148-149, Major bowel procedures

DRGs 82, 172, 173 - Respiratory neoplasams (82) and digestive malignancies - No post hospital care within 60 days.

DRG	Death	During Only	Before but No After	RRB	No B claims	60 Days +	Total
82	5	2	6	0	0	0	13
172	5	2	6	2	0	0	15
173	5	2	3	0	1	4	15

DRG 148-149 Major Bowel. Virtually all cases with surgery.

ATTACHMENT 2

RECOMMENDATIONS FOR THE HCFA SINGLE EPISODE ANALYSIS

1. Remove Railroad Retirement Board Beneficiaries

2. Remove beneficiaries who elected not to be eligible for Part B for the duration of calendar 1981. (In particular, include deaths but exclude terminations from the data file.)

3. Eliminate discharges in medical DRGs with significant hospital operating room charges on their MEDPAR records.

4. Examine and (possibly) eliminate discharges in medical DRGs with a surgical procedure indicator on their MEDPAR records.

5. Include all surgery concomitant with the hospital stay regardless of place of surgery.

6. Exclude all data from patients with December admissions, not just the December admission data.

7. Examine and eliminate patients with group prepayment plan (GPPP) Part B records.

July 15, 1984

Progress during the Fifth Month

The extensive series of "cleaning" steps that have been applied to the 100% South Carolina data for single stays in our parallel project has resulted in the development of an applicable series of steps for the one percent sample BSR - MEDPAR data. A description of these steps is attached. These cleansing steps are to be viewed as further refinements applicable to the General Estimating method for refining calculations of the values of MD-DRGs, Deliverable No. 3, and also will be applicable with further modification to Deliverable No. 4.

Additional work involving testing of the cleansing algorithm for single and multiple stays is under way. This algorithm to be used with multiple stays will be available shortly.

July 15, 1984

DELIVERABLE 3 (Addendum)
General Estimating Method for Refining Calculations
of the Values of MD-DRGs

Our review of the data has indicated that a sequence of algorithms must be developed to compute average total reimbursements to physicians for Medicare patients hospitalized for specific DRGs. These involve cleaning the data and censoring from the data observations which either have a high likelihood of having been misclassified into particular DRGs or where miscoded data would prevent proper estimation of average charges of particular DRGs. Elimination of observations with incomplete, complicated, or absent Part B coverage will also be attempted in the data cleaning process.

For the most part, misclassifications involve true surgical cases which have been labelled medical. Missing or miscoded data often appear to be the cause of surgical DRG cases with no surgery. Correcting for these possibilities will begin with records of beneficiaries who have had only one hospitalization during the year, and will be further restricted to hospitalizations which do not occur in January, February, November, or December.

Potentially misclassified stay records will be identified as those with a medical DRG discharge and a Bill Summary Record indicating one or more surgical services provided in an inpatient setting in a period concomitant with the hospital stay where the submitted charge for the surgical service was $\geq$ \$300. (BSR records for periods potentially concomitant with a hospital stay will be identified and tagged with a stay indicator, i.e., a variable indicating that there may have been a hospital stay associated with the service in question.)

Potentially miscoded records involve those patients with a discharge from a surgical DRG where there is no Bill Summary Record indicating at least one surgical service for $\geq$ \$300 provided in an inpatient setting in a period concomitant with the hospital stay.

Based upon the cleansed data, average charges will be computed for patients with single hospitalizations by DRG based solely on services provided in an inpatient setting in a period concomitant with the hospital stay. Averages will be computed by type of service for all single stays within a DRG and for only the recipients of each type of service by DRG.

Refinements to account for the effects of windows of varying lengths are being examined using the more detailed original Parts A and B data from South Carolina and Florida. A decision of which data to use for the comparisons of single admission beneficiaries and multiples is pending.

Attachment 1 describes the cleansing algorithm.

Algorithm for MEDPAR - BSR Data Preparation/Cleaning

I. Initial exclusions

Railroad Retirement Board Beneficiaries

ESRD

Non-95'ers

1980 admissions

1982 discharges

(Exclude those not eligible for B full 12 months)

(GPPP enrollees)

II. Select all MEDPAR records in jurisdiction of interest (South Carolina only for the present time)

Exclude beneficiaries discharged from other than general short term hospitals

Exclude beneficiaries discharged with DRGs 468-470

III. Partition MEDPAR records into 2 groups: singles and multi's. Sort by HIC numbers.

IV. Partition singles to select beneficiaries who have admissions or discharges in January, February, November or December

Hold these in File 1

V. Match and fetch and sort BSR records for single's file and multi's MEDPAR files

Sort by HIC, "from" month, "to" month, place of service, type of service

VI. Tag BSR records

Convert admission date and discharge date on MEDPAR record to admission month and discharge month

Set a stay indicator on all BSR records which span a MEDPAR stay month, i.e., if the BSR "from" or "to" months or any month in between match either the hospital admission or discharge month, then a non-zero stay indicator variable will be added to the BSR record.

VII. Cleanse (and count) singles

Hold in File 2 all singles where there is a BSR record with a stay indicator but there is no spanned hospital stay.

Hold in File 3 all singles where: DRG is medical and there is a BSR record with (a) stay indicator, (b) POS = inpatient and (c) TOS = surgery.

Hold in File 4 all single where there is no BSR record for inpatient service and a stay indicator.

VIII. Print cleansed singles in HIC order [with MEDPAR record(s) followed by BSR record(s)]

IX. Print the multi's in HIC order with MEDPAR record(s) followed by BSR record(s) for development of additional cleansing algorithms.

July 23, 1984

Transmitted herewith is our first set of suggestions for algorithms for apportioning values of physicians services in cases of multiple admissions.

We have begun the process of validating these, using a combination of the one percent MEDPAR-BSR data and a one percent matching sample of the South Carolina Blue Cross-Blue Shield data. We anticipate that we will be able to suggest choices among these suggested algorithms based on these efforts.

Algorithm for Apportioning Values of Physician Services in Cases of Multiple Admissions

Given the availability of BSR and MEDPAR data from all the states, the opportunity clearly exists to estimate national physician service values associated with hospital stays. The BSR date information is month specific rather than day specific, however, and a significant number of records cover more than one month — some may cover services for an entire year. This may be particularly troublesome given that just over half of all Medicare admissions in any one year will be part of a multiple of two or more stays during that year. Thus there is a potential need for an algorithm to allocate physician service values for records which appear to span the stay dates of two or more hospital stays.

Although more than half the stays in question will be multiples, not all multiple stays will require an algorithm that involves more than simple assignment by dates of service. Based on an examination of records from 229 South Carolina multiple admission beneficiaries with 550 stays, fifty-six percent had BSR records for inpatient services that could be unambiguously associated with specific hospital stays. When broken down by numbers of admissions, these records show the following:

# Stays	Total Beneficiaries	Allocable Beneficiaries	% of Total
2	143	98	68
3	57	26	46
4	17	4	24
5	10	1	20
6	1	0	0
8	1	0	0
Total	229	129	56

(In fact, these data suggest a 68% progression as additional stays are added. 68% * 68% = 46%, 68% * 46% = 37%, 68% * 31% = 21%.)

The reason that many BSR records will be allocable by date is that most records have only a single month of service coverage and most (56%) interstay gaps will be greater than one month. The most likely situations where allocation algorithms will be required involve interstay gaps of two weeks or less. This is approximately 30% of the total.

If the national statistics were comparable to the South Carolina experience, just more than three-quarters of all stays might be available to estimate inpatient physician service values based on dates of service. Unfortunately, the remaining quarter may still require clarification. The multiple stays with short interstay gaps may be just those cases of seriously ill patients that have significantly greater physician service values. Hence there remains a need to examine allocation algorithms.

Possible algorithms include the following (all assume that if the "from" month "to" month dates match only one stay that record will be allocated to that stay):

1.1 Allocate physician service values equally between stays.
Example:

If a BSR record potentially spans only two stays, allocate half to each; if three stays, allocate 1/3 to each of three stays, etc.

1.2 Allocate proportionately by spanned lengths of stay.
Example:

Given a BSR record for June and two stays that include 2 days in June and 3 days in June, respectively, allocate 40 percent to the first and 60 percent to the second.

1.3 Allocate to the stay with the greatest spanned length of stay or default to algorithm #1.
Example:

Same as example 2. One would allocate all of the BSR values to the second stay. (If both stays had equal spanned lengths of stay, allocate half to each.)

1.4 Allocate stochastically between spanned stays based on the relative spanned lengths of stay.
Example:

Same as example 2. One would use a random number generator or some other device that would allocate the record to the first stay 40% of the time, to the second stay 60% of the time. This algorithm would be superior to #3 for those cases with more than one BSR record that required allocation.

(1.1S-1.4S Identical to 1.1-1.4 but for BSR records with surgery as the type of service — and allowed charges per service > $300 — only consider stays with surgical DRGs. If no such stays are in the decision set, eliminate that beneficiary's records. A comparable rule could be applied to anesthesia records.)

1.5 If the type of service is medical, allocate the record to that stay where the difference between the length of stay and the number of services is one (or less). If there is more than one such stay or no stay meeting this criterion, default to 1.4. (Examination of the detailed Part B records for South Carolina indicates that a very common pattern of medical practice results in a bill with a hospital visit on each day the patient is in the hospital. Thus the number of services will be equal to the length of stay plus one.)

These algorithms have the advantage of involving only one pass at the data. That is, each BSR record need be read only once to determine its allocation. Another set of alternatives could involve two or more passes through the data. In particular, unambiguous BSR records would all be allocated and decisions on the remaining records, if any, would include consideration of the initial allocations. For example, consider a case where there were two surgical DRG discharges and two surgery BSR records to be allocated, only one of which (BSR records) spanned both hospital stays. The unambiguous BSR record might be allocated in the first pass. The ambiguous BSR record might then be allocated to the stay which did not initially receive a surgery record.

Multiple pass algorithms appear to take better advantage of available data, but they would involve additional cost. Further examination of such algorithms will be deferred pending review of results from single pass algorithms for selected sample cases.

An issue which complicates both multiple and single pass algorithms for inpatient services arises in consideration of data censoring/cleansing rules. For example, for single admission medical discharges, a cleansing rule has been proposed to delete cases with spanning inpatient surgical charges of $250 or more. Should an algorithm be adopted that would allocate surgical charges only to surgical DRGs, a case where the surgery for a discharge which was both a) miscoded as medical and b) paired with a subsequent discharge for a surgical DRG might escape later censoring.

Partly for this reason, the following cleansing rules are proposed:

I. Partition multiples to select beneficiaries who have admissions or discharges in January or December. Hold these records in file M.1.

II. Match and fetch sorted BSR records. (Sorted by HIC, "from" month, "to" month, place of service, and type of service.

III. Tag BSR records

Convert admission date and discharge date on MEDPAR record to admission month and discharge month. Identify all months of hospitalization. Set a stay indicator on all BSR records which span one or more MEDPAR stay months.

IV. Cleanse (and count) multiples

Hold in file M.2 all multiples where:

A. A BSR record indicates place of service is hospital but no stay indicator has been added, possibly suggesting a missing A record.

B. A BSR record indicates inpatient surgery with an allowed charge per service > $300, but there is no spanned hospital stay with a surgical DRG.

C. There is a hospital stay for which there are no potentially matched BSR records.

Comparing Inhospital Physician Service Values of Beneficiaries with Multiple Admissions

In addition to examining windows associated with multiple admissions, the inhospital physician services should be examined to ascertain whether admissions which are part of multiples are significantly different from "single" admissions in the same DRG. For the purpose of conducting this comparison, DRG specific admissions within multiples will be partitioned into three groups: those admissions which are the first of a multiple within the calendar year, those admissions which are not the first of a multiple and which occur

within four months of a prior discharge during the year, and those admissions which are not the first of a multiple and for which the interstay gap of the immediate prior admissions is more than four months.

August 22, 1984

Transmitted herewith is a Draft Plan for simulating the effects of a physician DRG payment policy.

Alternative I corresponds to the concept of payment to a single, lead physician.

Alternative II prorates physician charges.

August 22, 1984

Draft plan for simulating the effects of a physician DRG payment policy

ALTERNATIVE I

STEP 1 -- For each hospitalization, identify a "lead physician"

For medical DRGs, the "lead physician" will be identified as that provider with the largest number of medical services (type of service, medicine) provided in-hospital spanning the hospital stay.

For surgical DRGs, the "lead physician" will be identified as that provider with the largest proportion of allowed charges for surgical services (type of service, surgery) provided in-hospital spanning the hospital stay.

Cases where a lead physician cannot be identified will be counted but partitioned away from the rest of the exercise.

STEP 2 -- For each hospitalization, compute the total allowed charges and the percentage of total allowed charges attributable to the "lead physician" for each window of interest.

STEP 3 -- For each jurisdiction and each window of interest, determine a physician DRG payment level.

STEP 4 -- For each hospitalization compute:

The difference between total allowed charges and the relevant DRG payment level.

The difference between total allowed charges attributable to the "lead physician" and the relevant DRG payment level.

STEP 5 -- Aggregage these differences by provider and examine the distribution of positive and negative differences

By specialty of the "lead physician"

By jurisdiction

By DRG

By window of interest

August 22, 1984

Draft plan for simulating the effects of a physician DRG payment policy (Continued)

ALTERNATIVE II

STEP 1 - For each hospitalization compute the total allowed charges and the percentage of total allowed charges attributable to each physician for each window of interest.

STEP 2 - For each jurisdiction and each window of interest, determine a physician DRG payment level.

STEP 3 - Multiply (prorate) the DRG payment level by the percentage for each physician for each window of interest.

STEP 4 - For each hospitalization compute:

The difference between total allowed charges and the relevant DRG payment level

The difference between total allowed charges attributable to each physician and the relevant prorated DRG payment level.

STEP 5 - Aggregate these differences by provider and examine the distribution of positive and negative differences

By specialty of each physician

By hospital provider number

By jurisdiction

By DRG

By window of interest

October 22, 1984

Since our last written progress report, we have developed a computer program to count and display the MEDPAR, Bill Summary Records, and OPD records for beneficiaries in the 1% sample. One aspect of this effort has been focused on identifying those patients who have ambiguous, inpatient BSRs, i.e., those which span two or more hospital stays. An interesting finding we have labelled the "68% rule." Specifically, 68% of the beneficiaries with the two stays only have no ambiguous inpatient BSRs $(68\%)^2$ or 46% of beneficiaries with 3 stays only have no ambiguous BSRs, $(68\%)^3$ of those with 4 stays only, and so on. Given that multiples of 2 represent roughly half the multiple stays, 3's one quarter, 4's one eighth, etc., the 68% rule imples that just over 50% of the multiple stays will have no ambiguous claims. In the actual counts for Florida, 48% of multiple stays did not involve any ambiguous claims. In fact, 50% may be an underestimate of the number of usable BSRs for inpatients since a beneficiary may have both ambiguous and unambiguous records.

During September and October, these approaches to manipulating the multiples data have been conveyed to the HCFA staff conducting the intramural work in meetings with Ira Burney, Marilyn Newton, Sherry Terrell and Stephens Jencks, among others. In particular, the cleaning algorithms used for the singles have been adapted to handle the multiples using the South Carolina intermediary and carrier data. These adaptations have been described to HCFA staff to assist in developing comparable algorithms for the intramural study.

Finally, the programming for data displays for the multiples' MEDPAR, BSR, and OPD data are being modified to assess potential allocation algorithms for the roughly 50% of the patients who do have ambiguous claims.

November 22, 1984

Since our last progress report, we have been trying to address the problem of appropriate methods for handling records of beneficiaries with multiple hospitalizations in a single year. Two minor new problems have been identified.

The first involves the presence of duplicate Part A records in the South Carolina intermediary's history file. Since we had not anticipated this possibility, beneficiaries with duplicate copies of a single Part A record were classified as patients with multiple stays because they had more than one Part A record. For the purpose of proceeding with the data manipulation, we have treated these multiple claims as if they were transfers. As with transfers, we have eliminated both claims. This resulted in a net additional loss of 803 stays in South Carolina, leaving the final "cleaned" multiple stay data set with 35,668 observations.

The second problem involves apparently misdated Part B claims for services provided in an inpatient setting but where the date of service lies between two hospital stays which are less than 10 days apart. The experience with the singles data suggested that these claims might best be assigned to the second of the two stays in question. Producing a computer program that implemented this algorithm reliably has proved to be more complicated than anticipated.

The recent experience with these new problems with the multiples suggests the utility of restating potential problem areas that can be anticipated in handling the in-house HCFA data:

MEDPAR and/or other Part A data

- Duplicate claims or -- worse -- near duplicate claims.

- Transfer episodes. These obviously involve at least two institutions, and they often involve two or more different DRGs.

Part A merged with Part B data (Bill Summary Records or BMAD)

- Part A records with no matched Part B records for the dates or place. Some of these will involve beneficiaries not eligible for Part B; some will involve missing Part B claims; some will involve miscoded Part B claims.

- Part B records apparently for inpatient services which do not match stay dates of any Part A records.

- Part B records which indicate major surgery for patients discharged in medical DRGs. Pre-1983 Part A records may contain miscoded surgical information or they may omit surgical coding altogether.

Our recommendation is to delete such observations from any data set to be used in examining the physician DRG question. To date we have no evidence that these types of data problems are associated with either high or low charges for either Part A or Part B claims.

January 22, 1985

Since our last progress report, we have prepared a variety of summary documents and presentations of the data from this study.

You have received 10 copies of the "Final Report." George Greenberg has received 15 copies of that report and will be distributing copies to six members of the advisory committee. You should also have received 2 copies of the tables containing all 470 DRGs. This is a separate document which will not normally be made available with the Final Report.

The Executive Summary for the report is based on the presentation to the Project HOPE Conference, January 16-17. Howard West and Peter McMenamin attended the conference on both days, and Peter also attended the Prospective Payment System Technical Advisory Panel meeting on the night of the 17th.

Enclosed are tables with the results of the 8-month/12-month and single/multiple comparisons for Part A charges. Tables for the Part B charges were included in the Final Report.

The general results of these comparisons parallel the results from the Part B work. One concludes that there are no substantial differences between the 8-month and 12-month samples. However, the sample of single stays is not a good representation of the multiple stays, and vice versa.

There is a very slight association between extreme t values for Parts A and B in the 8/12 comparison for individual DRGs. The DRG specific association is more obvious in the single/multiple comparisons, particularly for the most extreme t values. This may be due to tails in both the Part A and Part B charge data.

8-month/12-Month Comparisons: Part A Charges and Part B Charges

DRG	PART A Diff.	PART A t	PART B Diff.	PART B t
12	89	.33	13	.48
14	68	.47	-7	-.64
15	13	.11	-3	-.20
17	40	.18	-6	-.29
24	-103	-.46	-10	-.38
39	-14	-.63	-9	-.82
40	-50	-.21	-8	-.18
65	8	.06	7	.39
68	169	1.07	-9	-.45
75	390	.73	127	1.28
82	97	.27	13	.28
87	452	.64	45	1.15
88	101	.58	-1	-.08
89	156	1.22	6	.67
96	-63	-.35	-11	-.85
101	338	1.14	4	.15
107	-585	-.69	-50	-.20
116	14	.04	-7	-.12
122	101	.83	13	1.30
123	-14	-.04	-16	-.67
127	111	.78	8	.72
128	153	1.00	9	.45
130	68	.32	15	.63
132	88	1.19	7	.88
134	112	1.30	9	1.00
138	69	.70	0	0.0
140	44	.76	1	.11
141	29	.27	-2	-.13
143	40	.58	4	.36
148	541	.90	-5	-.08
154	342	.65	24	.33
157	115	.76	0	0.0
161	-17	-.09	-5	-.23
172	-221	-.39	-34	-.79
174	95	.40	14	.93
182	-13	-.28	-1	-.17
188	55	.22	-6	-.23
197	40	.17	0	0.0
204	191	.35	-7	-.21
207	71	.42	8	.50

8-month/12-Month Comparisons: Part A Charges and Part B Charges

DRG	PART A Diff.	PART A t	PART B Diff.	PART B t
209	-6	-.03	26	.41
210	59	.34	7	.29
225	4	.04	15	.45
233	242	.83	77	.68
236	465	1.09	2	.05
243	38	.58	2	.20
244	49	.45	25	1.19
253	-176	-.74	-21	-.70
262	8	.11	-7	-.32
269	46	.14	25	.78
277	-90	-.44	-9	-.56
280	211	1.42	4	.27
294	94	1.52	5	.71
296	95	.56	-10	-.56
316	-393	-.61	-31	-.48
320	112	.94	14	1.16
323	-80	-.56	-7	-.35
336	58	.65	5	.24
356	3	.02	20	.61
364	18	.20	11	.50
395	28	.15	-3	-.18
403	-4	-.01	-10	-.20
413	-54	-.11	8	.16
426	124	.98	12	.44
429	166	.69	-4	-.14
430	-25	-.11	-4	-.10
438	-37	-.14	-6	-.24

Single/Multiple Comparisons: Part A Charges and Part B Charges

DRG	PART A Diff.	PART A t	PART B Diff.	PART B t
12	66	.23	+4	.15
14	261	1.68	+19	1.36
15	-73	-.74	-40	-2.86
17	231	.79	-18	-.75
24	196	.86	-21	-.84
39	-32	-1.33	-29	-2.07
40	201	.73	+66	1.40
65	175	1.23	+44	2.00
68	339	2.28	-4	-.19
75	866	1.46	+220	2.31
82	-301	-.97	-83	-2.13
87	-393	-.61	+27	.75
88	104	.66	+6	.50
89	521	3.92	+30	3.00
96	-43	-.26	-7	-.54
101	546	1.78	-10	-.40
107	-578	-.68	-125	-.53
116	405	.97	+17	.28
122	252	2.02	+23	2.09
123	-488	-1.41	-26	-.84
127	-107	-.84	0	0.0
128	124	.91	-10	-.59
130	90	.44	+8	.36
132	203	2.74	+14	1.75
134	501	4.36	+46	4.18
138	134	1.34	-2	-.18
140	115	1.98	+6	.67
141	200	1.52	+10	.53
143	148	1.85	+14	1.07
148	710	1.19	-108	-1.74
154	451	.90	+95	1.30
157	373	2.10	+43	.93
161	-52	-.28	-34	-1.42
172	-726	-1.51	-23	-.62
174	24	.11	+32	1.88
182	198	3.47	+16	2.29
188	58	.26	-19	-.76
197	566	1.84	+47	1.24
204	-122	-.29	-24	-.80
207	212	1.14	+20	1.11

Single/Multiple Comparisons: Part A Charges and Part B Charges

DRG	PART A Diff.	PART A t	PART B Diff.	PART B t
209	173	.60	+79	1.01
210	201	.99	+32	1.03
225	294	1.26	-146	-3.65
233	1196	2.41	+132	1.20
236	-392	-1.06	-72	-1.76
243	139	1.90	-3	-.30
244	57	.45	-10	-.59
253	-18	-.06	-31	-.97
262	280	1.85	+58	1.26
269	463	1.18	+58	1.14
277	253	1.17	+23	1.28
280	434	2.26	+31	1.48
294	283	4.22	+24	3.00
296	310	1.78	-12	-.71
316	-56	-.09	+.14	.24
320	342	2.83	+26	2.17
323	-88	-.64	-26	-1.24
336	85	.71	-39	-1.86
356	174	1.11	+29	.66
364	-15	-.16	+7	.33
395	-119	-.70	-54	-3.38
403	-151	-.31	-42	-.95
413	-113	-.28	+33	.72
426	248	2.02	+47	1.74
429	55	.28	-8	-.29
430	-174	-.91	-7	-.20
438	-667	-2.89	-40	-1.82

APPENDIX I

EXECUTIVE SUMMARY

EXECUTIVE SUMMARY: SOME LESSONS RELATING PHYSICIAN PAYMENTS TO DRGS

This study was initiated primarily to assess the potential methodologies that might be employed in examining the merger of Part A and Part B data rather than explicitly to develop methods that might be employed in establishing a physician payment system based on the DRGs. However, given the immediacy of the requirement that the Department of Health and Human Services (HHS) report to Congress on the feasibility and advisability of physician payments using the hospital DRG system, this summary will be focused on the lessons that have been learned in the course of the project as they relate to a potential physician DRG system.

Consideration of certain aspects of such a system was unavoidable in developing the programming for tables and other displays of the data -- especially given the use of the DRG classifications to partition and organize both the Part A and Part B data. As a result, certain somewhat tentative conclusions have been drawn from the exercise with respect to the problems that might be encountered in trying to implement a physician DRG system. Because of the nature of the work, most of these lessons are relevant to the use of data to develop such a system. As might be expected, the statistics generated in exploring these methodologies are also relevant to the development of a variety of possible reimbursement system parameters.

With respect to the "ultimate" questions of the feasibility and advisability of a physician reimbursement system related to the current DRGs for hospital payment, the data available offer only partial answers. From a strictly conservative perspective based on the analysis of the merged carrier and intermediary data from two states, one would conclude that a physician DRG system is <u>not inconceivable</u>. A few components necessary for the development of such a system are feasible, but the data at hand do not reveal whether such a system would be advisable.

The data used in the study came from intermediary and carrier files from the states of South Carolina and Florida for calendar year 1981. HCFA abstract data for the same states for that year from the MEDPAR and Bill Summary Record Systems were also analyzed. Basically, the carrier and intermediary data (and

the directly derived 1981 HCFA in-house data) are most relevant to the questions associated with establishing rates for MD DRGs.

We have presumed that a process similar to the rate development process used in the hospital reimbursement system would be employed, perhaps with some variations. (This involved establishing, in effect, a hospital cost relative value scale for DRGs based on the arithmetic averages across observations after deleting not quite one percent of the data as outliers.) For this reason, this study has focused on the potential biases that might be present in the alternative algorithms to estimate average allowed physician charges by DRG using the available data. There are several issues that we believe have been identified and clarified, if not resolved, by the Mandex analyses. These include:

- merger of Part A and Part B data
- completeness of the merged data
- errors in the data
- data filtering
- hospital PPS multipliers
- variations in physician services
- variations in allowed charges
- variations in physician coordination
- outliers
- single versus multiple hospitalizations

PART A PART B DATA MERGER

With respect to the question of whether these data can be merged to identify physician charges associated with Medicare beneficiaries' hospital episodes for specific DRGs, the answer is yes. The rates of failure to match beneficiaries' Part A and Part B records were 1.9 percent in South Carolina and 2.9 percent in Florida. Both of these rates are well within the realm of expectations given that HCFA statistics for 1981 indicated that 3.3 percent of beneficiaries in South Carolina and 2.1 percent in Florida had Part A but not Part B coverage. One would also expect that the non-matches might reflect the experience of beneficiaries in transit in that Part A claims for "travelers" would be directly initiated by the hospitals, but non-assigned Part B claims might be misdirected by beneficiaries resulting in incomplete files for those beneficiaries. Individual carriers' data will appear to be deficient with respect to physician practices that are interstate, or that involve inter carrier jurisdiction. This may

have affected the Florida results because a different carrier handled the claims from physicians located in the Miami area (Dade and Monroe counties) in 1981.

DATA COMPLETENESS

The issue of data completeness must be addressed from two different perspectives, namely, are the data complete with respect to services that are billed/reimbursed and are the data complete with respect to services that are provided? There is no inexpensive way to assess whether records of all physician bills are contained in carrier files. Our belief is that if the carrier received a bill from a physician for services rendered to a Medicare beneficiary that bill will be reflected in the data. Further, Mandex has reviewed and compared HCFA Bill Summary Record data with the "originating" data from South Carolina Blue Cross Blue Shield. The concordance between these two data sets is extremely high, suggesting that with respect to the extent of available data, there is almost no (significant) deterioration that conceivably might have been introduced in the data abstraction process that creates the Bill Summary Records.

The question of apparently "missing" bills is a complicated one. The following scenarios are all plausible yet indistinguishable -- if even discernable -- in the data: (1) a Medicare patient receives physician services, a non-assigned bill is rendered, but the beneficiary fails to file a claim. Or (2), an assigned claim is filed but not included by the carrier in the beneficiary records. Or (3), the physician services are rendered but without a bill. Or (4), the Medicare beneficiary does not receive a particular physician service that might be typical of the care received by comparable patients in the particular DRG in question. In all cases, one or more typical services may appear to be missing.

We believe that for the bulk of physician services rendered to Medicare beneficiaries, there will be a bill and hence a record in the files. However, it is our impression that the billing data are inconsistent with the conventional wisdom with respect to what would be perceived as desirable patterns of physician continuity of care. Specifically, they suggest a lack of continuing

care for a substantial portion of these patients. The review of data with respect to "windows"[1] of treatment surrounding a hospitalization indicates that a significant number of Medicare patients do not receive care either preceding their admission or following their discharge. This phenomenon appears to occur a sufficient number of times to suggest that, although there may be a carrier data problem of misplaced claims, it is more likely that the claims simply are not being filed with the frequency that might be expected.

If this pattern were only apparent with respect to the windows, it might not significantly affect the deliberations with respect to a physician DRG system. (In fact, given the verification of this pattern of paneless windows in both Florida and South Carolina, the focus of the study shifted to strictly inpatient services and those procedures with dates of service coincident with hospital stay dates.) Even on the inpatient side, however, there are enough examples of cases with apparently missing bills for physician inpatient services that an initial DRG reimbursement policy planning dilemma becomes apparent. The dilemma consists of the choice among the alternative sets of observations to be included in the computation of average allowed charges.

Whether these "missing" services are services that have been rendered but are lacking bills or are services that have simply not been rendered, their absence under the Medicare claims data system means that Medicare reimbursements are lower than they might have been otherwise. Because those bills which are in the files are consistent with existing total levels of Medicare reimbursements, any attempt to delete cases which fail to meet plausible quasi-clinical standards of completeness might well imply an increase in Medicare outlays.

Specifically, those patients whose care as reflected in billings does not meet standards with respect to hospital visit frequency, for example, may not be receiving the best quality care or are failing to present claims for services actually provided. In either case, Medicare is not being billed for the additional services that might be necessary to bring that care up to quality standards -- even if only with respect to documentation.

---

[1] Windows in this study are defined as periods of time 7, 15, 30, 45, and 60 days before hospital admission and after hospital discharge.

If there were five such cases in a DRG with 100 observations, one would have the choice of computing average allowed charges based on either the entire group of 100 or only the group of 95 that met the standard. If all 100 observations are included in the computation of the rates, the resulting physician DRG system rates for that DRG might be budget neutral with respect to the current physician reimbursement system, but the averages by DRG might be less than are currently reimbursed for most episodes of care in that DRG. Excluding those "non-standard" observations might imply sacrificing budget neutrality, indeed increasing reimbursements if all 100 cases were presumed to be reimbursed at the level of average allowed charges from the group of 95 "standard" observations. Alternatively, one might exclude the apparently below standard observations but only to develop a physician DRG relative value scale. For rate-setting purposes, all observations would be included. Budget neutrality for the entire set of DRGs would then imply an across-the-board reduction in all prospective physician DRG payment rates. That is, the average allowed charges estimated for the relative value scale would be converted into rates using a multiplier with a value less than one. This would affect even those DRGs (perhaps by states or physician specialties) where there had been no observations of "non-standard" care.

To recapitulate, the Mandex review of the data in the files suggests that those data are in general unremarkable. We do not believe that an appreciable number of claims have been lost by the carriers in question, but in a substantial number of cases some expected claims appear to be missing. Hence although there may be a data problem, we do not believe this results from a data processing problem. In that sense, the data are believable even if some observations appear to be implausible.

ERRORS IN THE DATA

Prior to the imposition of the hospital prospective payment system (PPS), the primary information of importance on Medicare hospital claims involved the beneficiary identifier, the hospital identifier, and the hospital charges. Neither diagnosis nor procedure, if any, made a difference, even though the claims forms

allowed for that information to be transmitted to the Medicare intermediaries. For that reason, those items were unlikely to be entered with any deliberate bias although they may not have been carefully entered.

The Mandex review of the merged Part A and Part B data in both Florida and South Carolina has revealed a significant problem of apparently inadequately coded Part A data, particularly with respect to the surgeries. We have observed many cases where the hospital claim record does not indicate the presence of a surgery while one or more Part B claims directly indicate a surgery. Although this situation will improve given the importance of surgery codes to reimbursement under PPS, data available or that become available with respect to any quarter preceding October 1983 are likely to suffer from missing or inadequately coded surgeries.

For the most part, the effects of these miscodings are unambiguous. The grouper program searches for specific surgery codes once a particular diagnosis code is recognized. That program will not match an omitted code, a code which is incomplete, or one which has otherwise been miscoded. And in those cases where two or more surgical codes might have been entered on a claim by a hospital, choice of a relatively insignificant surgery as the first listed on the claim might also lead to a grouper failure to match. Failure to match in all cases results in the claim being assigned to a Medical rather than a Surgical DRG. Except for a single case identified to date, Medical cases are never misclassified to Surgical DRGs.

The potential effect of this problem for implementing a physician DRG system is the possibility of overestimating the average allowed charges that have been associated with hospitalizations in Medical DRGs. (There is also the possibility that the weights for hospital payment purposes have been overestimated for the Medical DRGs. It does not appear that the surgery average allowed charges — or hospital weights -- have been underestimated.)

With respect to prospective rate setting and, possibly, establishing baseline data for evaluation purposes, two solutions exist for addressing this problem. One would be to correct the Part A data to the extent possible using

the Part B information. A second would be to filter the data with respect to suspect observations. Both approaches were considered by Mandex in this study, and to varying degrees, both were implemented during the study.

Part B data were used to correct Part A surgical codes which had been coded in machine readable form with only the first three of four ICD-9-CM surgical codes, as required for accurate assignments by the grouper program. Adding the correct fourth digit involved case-by-case review of observations by medical coding staff, and, given the disparities between ICD-9-CM and the HCFA common procedure coding system (HCPCS), not all cases could be resolved unambiguously.

Partly because of this experience and partly because of our access to files said to represent 100 percent of the data, we chose to use filtering as the primary means for cleaning the data. We believe this general approach is more economical than attempting to correct apparently suspect observations. There is no evidence that this approach will lead to biased estimates of average allowed charges by DRG, since we found no association between midcoded surgeries, for example, and either relatively high or low Part A charges.

OTHER DATA FILTERING

There were several types of problems encountered in the review of the data which were perceived as potentially troublesome, and which were resolved by data filtering. Some of these deletions would not necessarily apply to a more general physician DRG program, but were incorporated into the current study to retain a tighter focus on the scope of work. For example, hospital stays in non-short term general institutions were not examined. These institutions are not currently covered by the hospital prospective payment system, and a special study to examine the potential applicability of a PPS system for those institutions has been mandated. The records of the relatively few beneficiaries with End Stage Renal Disease appearing in the data set were eliminated since their patterns of physician service utilization were unlikely to be comparable to those of other beneficiaries given the use of capitation payments in that

program. Further, stays which were not "groupable" were eliminated from consideration in this study since they did not represent any particular set of diagnostic categories. Because the incidence of such cases should decline over time, a separate rate might not be needed for this category.

One other category filtered from the Mandex analyses involved hospital transfers. Nearly 9 percent of the Part A records available for the analysis of stays of beneficiaries with more than one hospitalizations were filtered out because they involved one or more transfers between hospitals. These episodes may require a special rate-setting approach because they often involve two apparently different DRGs. In addition, several cases were observed where a beneficiary was admitted to one hospital, transferred to a second, and then returned to the first. Ultimately, a method would have to be developed to incorporate these stays in any physician DRG system that might be implemented.

The two remaining categories of patient data filtered from the analyses involved beneficiaries whose matched Part A and Part B data indicated either (1) no physician inpatient services or (2) physician inpatient services apparently rendered while the beneficiary was not in any hospital. The first of these data filtering situations occurred in 2 percent of the stays in South Carolina with more than one stay and nearly 3 percent of the stays of beneficiaries with only a single stay during 1981. Although these might have been retained in the data along with the records of other beneficiaries with apparently missing claims records, several instances were identified where a beneficiary had Medicare Part B coverage for only part of the year, and was ineligible for reimbursement of physician services during the time of the hospitalization. Because it was unlikely that a hospitalized beneficiary would not have at least a single physician claim for the period of hospitalization, these observations were presumed to derive from partial year coverage, and hence were eliminated from further analysis. In establishing a data base for rate-setting purposes, a more direct process of filtering should be considered to eliminate from the data the records of beneficiaries without Part B coverage during the time of their hospitalization(s). Eliminating the records of all beneficiaries with

only part-year Part B coverage, however, might lead to bias due to the potential extraordinary initial utilization by beneficiaries in their first year of Part B coverage.

Observations of apparently missing Part A records represented about 16 percent of the stays of South Carolina beneficiaries with multiple stays in 1981 and 8 percent of the single stay beneficiaries. (The comparable statistics for Florida were 15 percent and 8 percent, respectively.) Although some fraction of these observations indeed represented cases of missing Part A records, most involved errors in the dates of service on the Part B bills. (It is not known whether these are errors in the claims or are data entry errors.) Should a DRG-related physician reimbursement program be constrained to cover only those physician services provided during the time of hospitalization, misdated claims have the potential to bias downward estimated average allowed charges by DRG, since those claims might not be properly allocated to the hospital episode in question. However, because such claims have been reimbursed in the past, since the carriers, for the most part, have not had the requirement much less the means to verify hospitalization dates, elimination of such patient observations from a rate-setting data base should not bias the estimates.

In examining cases of apparently misdated claims, it was discovered that many cases involved relatively minor misdating. For example, it was often observed that a claim for an EKG shown as being taken in the hospital and probably performed upon the beneficiary's admission had a date of service one or two days prior to the date of admission. Since there were a sufficient number of such cases, the inclusive period of hospitalization was "extended" for allocating claims for physician inpatient services to a stay. Specifically, a five-day tolerance on either side of the stay was adopted to correct for these minor errors in dates of service.

Apparent data entry errors also occurred with regard to place of service. Claims having a date of service that fell within the hospital stay dates but had a place of service code other than inpatient were also observed. These cases were included as though place of service had been given as inhospital.

PPS MULTIPLIERS

Given the ongoing implementation of the hospital PPS system for Medicare, the possibilities of tying a potential DRG-related physician reimbursement system to that hospital system warrant exploration. In particular, one or more multipliers that would easily convert hospital rates to physician rates would have a great appeal. There are some obvious drawbacks to such an approach. For example, there is the need to periodically adjust the multipliers given the different rates of growth in unit prices between hospital and physician services. However, if feasible, not having to develop and explain a DRG-related physician reimbursement system that would be distinct from the hospital DRG based PPS system would greatly reduce the relative administrative burdens of implementation.

Mandex examined both a single multiplier for all DRGs and a set of DRG-specific multipliers with one for each DRG. We find no empirical support in any of the data for any kind of multipliers.

Tables of the ratio of physician-billed charges to hospital charges were prepared to assess these possibilities. It was known that, nationally, hospital reimbursements were roughly 70 percent of charges in 1981; physician allowed charges were approximately 77 percent of billed charges during that year. With respect to a single multiplier, there was no uniformity in the ratios in either of the two states. Surgical DRGs tended to have higher ratios than Medical DRGs. However, there was no uniformity in ratios within either of these categories.

With respect to DRG specific multipliers, there may be somewhat greater uniformity, but a different problem immediately became obvious in comparing these ratios between the two states. Florida physician charges relative to hospital charges are much higher than those of South Carolina physicians, as well as appearing to be more variable. A single set of multipliers based on the Florida data, for example, could lead to windfall physician gains in South Carolina -- not to mention quantum jumps in beneficiary coinsurance burden in those cases. On the other hand, Florida physicians in general and/or beneficiaries with unassigned claims in particular might suffer windfall losses if the rates were

based on the South Carolina averages. A single set of multipliers based on national or regional data would cause similar windfall gains and losses.

VARIATIONS IN PHYSICIAN SERVICES

One presumption of a prospective rate system would be that by making reimbursement limits known in advance of the provision of treatment, provider incentives will simply encourage the efficient provision of care. With respect to hospitals, this may result in some small change in the frequency with which some ancillary services are provided and perhaps shifting to lower cost vendors of supplies. It is neither expected nor desired that there be dramatic changes in clinical practice for the purpose of increasing net revenues or reducing net losses with respect to patients of either hospitals or physicians.

The magnitude of the potential incentives for dramatic clinical changes is a function of the existing variations in treatments within DRGs. In the course of this study, Mandex has examined within-DRG variations with respect to (1) the specific physician services that were provided to Medicare beneficiaries in South Carolina in 1981 and (2) average allowed physician charges in both South Carolina and Florida during that year.

One initial objective of the study was to establish that a set of physician services common to all patients within a DRG could be identified as a starting point for examining variations in treatment within that DRG. The search for a set of common services, however, proved fruitless in all of the DRGs that were so examined. There was not a single specific HCPCS code for a physician service that could be observed among the claims of all patients in a particular DRG.

There are at least three phenomena that might make so uncommon a list of common services: missing claims, diversity of billing style, and clinical heterogeneity. An examination of these phenomena gives additional clues as to the feasibility and advisability of the reform under consideration.

The first reason why there might not be a list of services common to all patients in a particular DRG involves the apparently missing services discussed

above. There were observations in the data set in surgical DRGs, for example, where there was no physician bill for the surgery in question, even though there was a physician claim for anesthesia with a date of service identical to the date specified as the surgery date by the hospital. As indicated above, such instances do create a challenge to establishing prospective physician rates which maintain budget neutrality, but this is not an insurmountable challenge.

Unfortunately, eliminating or correcting such observations would not solve the problem of a lack of commonality in treatments within DRGs. There are too many instances observed involving the other two phenomena in question: diversity of billing style and clinical heterogeneity. Conceptually these are distinct phenomena although in practice the distinction may be a fine one. For example, one illustration of diversity of billing style might involve the choice between billing for a "brief hospital visit" versus a "limited hospital visit." Although there is also supposed to be a substantive as well as a taxonomic difference between those two types of bills, there are some physicians who never bill for a hospital visit of any intensity lower than "limited" regardless of how brief the actual visit may have been. Within the set of all medical DRGs, consolidation of visits into a smaller number of categories would allow identification of at least one physician service common to all patients.

Where clinical heterogeneity and billing style diversity begin to blend may be seen in some of the surgical DRGs. With respect to the clinical descriptors available in HCPCS for some of the major surgeries, there are often many fine clinical distinctions that can be recorded by the use of specific codes which relate to the conduct of the operation. Some physician practices may be relatively precise and avail themselves of the use of specific surgical HCPCS codes in all instances. Others may use a somewhat more generic billing style except in rare instances. Even though there may be a single generic operation common to all patients in a particular surgical DRG, the precise -- hence diverse -- billing style will make it appear that there are no common procedures.

At some level, the apparently small clinical differences that may or may not merely reflect billing preferences yield to substantial differences in medical

care. For example, within DRG 209, major joint operations, one can observe both total hip replacements and total knee replacements. Those are significantly different procedures with different demands on physician resources. (Average allowed charges for the knee operation, for example, appear to be higher than for the hip operation.) Within DRG 39, cataract excision records were observed either with or without codes claiming payment for the insertion of an intraocular lens. Those are different procedures. If a single reimbursement rate were paid for both, it is conceivable that there would be many fewer combined operations and many more readmissions solely for intraocular lens insertions.

Finally, there are a few surgical DRG "waste baskets" that may make sense for hospital services, but where a prima facie case might be made that there is little or no chance to observe clinical homogeneity. For example, DRG 233, "other musculo-skeletal and connective tissue operating room procedures" includes much of the orthopedic surgery that can be performed anywhere between a patient's fingers and his toes. It would be a surprise to find a single procedure that was common to all patients in this category. But for that reason alone, it would be questionable whether a single rate might be adequate to reimburse physicians for each of their patients in this DRG.

To recapitulate, there appear to be substantial variations in the patterns of service provided to Medicare beneficiaries even within individual DRGs. Although the potential reasons for such variations may involve both style and substance, they call into question the feasibility of developing meaningful single rates for physician services by DRG. If such a system could be found to be feasible, there would have to be sufficient uniformity in allowed charges to disregard the diversity in services.

VARIATIONS IN ALLOWED CHARGES

It is conceivable that, although there might be diversity in the types of procedures that are provided to patients in a particular DRG, the aggregate charges for those procedures might be more comparable. For this reason,

statistical measures of variation such as standard deviations, coefficients of variation, and R^2 estimates were computed using data for various subsets of the DRGs from South Carolina and/or Florida.

There is a great deal of consistency in the patterns observed in both of the two states, although this consistency does not imply that there is no variation in physician charges. The variances about some of the mean charges are substantial. Looking only at the 67 relatively high volume study DRGs in South Carolina, 53 have standard deviations between $127 and $300 with 22 between $200 and $300. Seven DRGs have standard deviations above $400. (Those DRGs with relatively high standard deviations tend to be surgical DRGs which also have relatively high mean charges. The means in fact, are relatively higher than the standard deviations.) The Florida distribution of standard deviations is comparable but the values are higher, as expected, given the higher levels of physician allowed charges in that state.

The relation between the standard deviations and the means is often summarized using a statistic called the coefficient of variation (CV), which is the ratio of the estimated standard deviation of a sample to its estimated mean. In the course of the study, Mandex computed coefficients of variation for all of the DRGs in both South Carolina and Florida, but the range of results derived from this exercise led us to more closely examine this statistic and the structure of the data underlying its computation.

As has been indicated elsewhere, this statistic is not robust with respect to skewed distributions, and the distributions of both physician and hospital charges are quite skewed. In fact, the major difference in variances between DRGs may be the presence or absence of outliers. Examination of a selection of moderate volume DRGs with a range of CV values revealed that the high values were primarily a function of the presence and distance from the mean of atypical observations. An examination of the normalized cumulative frequency distributions of these DRGs showed that the distributions of physician allowed charges were virtually the same in each DRG. Further, the distribution of hospital charges for these DRGs followed the same pattern.

That there are comparable patterns of variation is a useful finding -- if not quite a necessary condition -- to proceed with the examination of the possibility of a physician reimbursement system based on the DRGs. In this regard, there are two questions that need to be addressed. First, does the DRG classification system itself lead to a sufficient reduction in aggregate variance to warrant its use for physician reimbursement purposes? Second, given the use of that classification system, is there sufficient uniformity in physician allowed charges within DRGs to warrant its use for physician reimbursement purposes?

With respect to the first question, any classification scheme can be assessed in terms of the variance within the classified data compared to the variation observed in the original, unclassified data. The particular summary statistic that is computed is known as R^2. It can range in value between 0.0, no reduction compared to the original variance, to 1.0, a complete reduction.

The DRG system was designed using a variance reduction clustering technique which incorporated data on hospital diagnoses, surgical procedures, and charges. Since physician charge data were not included in the development process, since unlike hospitals there can be more than one physician involved in any hospital stay, and since on average there are nearly 100 physicians for each hospital, one might expect the hospital data to exhibit a greater reduction in variance than the physician data when classified using the DRGs. Still, if the R^2 statistics for the hospital and physician charge data were comparable, this might suggest there is sufficient uniformity in the physician charge data to resolve question number one.

Somewhat surprisingly, the R^2 for the physician charges consistently appears to exceed the hospital R^2. Using the South Carolina data for 1981, Mandex computed values of R^2 of .66 for physician charges compared to .26 for hospital charges. As a result, one might be led to infer that the DRG classification system designed for hospitals would work as well, if not better, for physicians. Examination of somewhat more disaggregated statistics suggests that these results are more of a statistical anomaly than clear-cut evidence on the feasibility of a DRG-based physician reimbursement system.

When the various data sets are partitioned between Surgical and Medical DRGs for the purpose of computing R^2 statistics, virtually the entire reduction of variance appeared to derive solely from the Surgical DRGs, even though these represented only about one-third of all discharges. In the South Carolina data, the Surgical DRGs produce an R^2 of .55 compared to an R^2 of .07 for Medicine. Aside from the apparent counterintuitive result that for physician charges the combined Medical and Surgical R^2 sometimes exceeded both the individual Medical and Surgical R^2's, the fact that the hospital data exhibited the same qualitative results prompted an examination of the computational basis of the R^2 statistic.

Basically, the R^2 statistic compares the variation about the arithmetic mean of a distribution to the sum of the estimated variations about each of the cell arithmetic means. The within cell variances are subtracted from the grand variance, and normalized by dividing by the grand variance. The smaller the variance with respect to the cells, the larger will be the R^2 statistic.

That within cell variances may persist despite the classifiction system has prommpted the popular notion that such variances are "unexplained" by the classification. Because the R^2 statistic is computed as the difference between the original variance observed in the data and the "unexplained" variance, it is often described as a measure of the variance explained by a classification.

Unfortunately, this nomenclature overstates the results of the process in question. Classification in and of itself "explains" very little. In fact, the R^2 computation involves a very basic comparison, i.e., how would the mistakes made by assigning to each observation the grand mean value compare to the mistakes that would be made by assigning to every observation the mean value of its cell? A set of R^2 statistics computed for two different classification schemes for the same data can suggest which of the two is better with respect to that data set. Comparing a set of R^2 statistics for the same classification scheme for each of two different data sets may only reveal for which of the two data sets would the use of the grand mean as an estimator be worse.

The R^2 results are unequivocal only with respect to the unadvisability of using the grand average of allowed physician charges per stay across all DRGs as the basis of payment for all hospitalized Medicare beneficiaries or the average across all Surgical DRGs as the basis of payment for all hospitalized Medicare beneficiaries with surgery. However, the results are not unequivocal with respect to the specific use of the alternative, viz., a payment system based on the average allowed charges by DRG. In particular, for the nearly two-thirds of the cases involving Medical DRGs, payment at the DRG specific average would make little improvement over payment at the average of all Medical DRGs.

If the answer to the question of aggregate variance reduction is unresolved, a positive answer to the second question with respect to relatively low within-DRG variance might bear on the question of the advisability of the use of the hospital DRGs for physician reimbursement purposes. Unfortunately, here again the data are not unequivocal.

As indicated, for the 67 relatively high volume DRGs studied in South Carolina, 31 exhibited standard deviations of less than $200. One might infer that nearly half of the DRGs had fairly tight distributions of physician allowed charges. Just by sample size alone, however, the estimates of standard deviations for the higher volume DRGs will tend to be more precise. In South Carolina in 1981, there were 279 DRGs in which there were at least 10 beneficiaries with a single hospitalization during that year. Of those DRGs, 108 exhibited standard deviations of less than $200, less than 40 percent.

What can be inferred from these measures of variance, i.e., what does a specific estimate of a standard deviation imply? Given the known statistical properties of standard deviations one can use those estimates to illustrate the potential consequences that might have been observed if the Medicare payment policy in 1981 had been to pay the arithmetic average allowed charge for all discharges within a specific DRG in lieu of the actual individual allowed charges as determined by the carriers.

The standard deviation across all observations in South Carolina for single stay patients was $203. (In Florida, it was more than double, $417.) The

table below is reproduced to demonstrate the distribution of payment disparities that might be expected, given a standard deviation of $200, had all reimbursements been made at the mean allowed charge rather than the individual observed allowed charge.

One would expect that in roughly 10 percent of the cases the disparity between the average allowed charge and the observed individual allowed charge would be $25 or less. That seems a tolerable error. Five percent of the cases would involve losses of $25 or less, five percent would involve comparable gains. In an additional 9 percent of the cases, the maximum expected disparity would increase to $50. As a result, not quite 20 percent of the cases would be expected to be in error by less than $50, half gaining and half losing. Although $25 to $50 "errors" may not be very high, the obvious additional implication is that in more than 80 percent of the cases, the "errors" would be expected to be larger. Indeed, in 62 percent one would expect to observe a payment disparity of no less than $100.

Implications of a $200 Estimated Standard Deviation: Estimated Percentages of Cases with Specific Ranges of Dollar Errors Associated with an Arithmetic Average Allowed Charge (MD DRG) Payment Policy -- Medical DRGs only

Range of Dollar Errors	Percentage of Cases
- 25 to + 25	10%
- 50 to + 50	19%
- 75 to + 75	27%
-100 to +100	38%
-125 to +125	46%
-150 to +150	54%

Such payment disparities will be even greater for the Surgical DRGs. Twenty of the sixty-seven high volume DRGs more closely examined in this study were Surgical DRGs. Of those 20, only three exhibited standard deviations less than $200 in South Carolina. Eleven of those DRGs had standard deviations in excess of $300. Although the Surgical DRGs may appear to be slightly less variable than Medical DRGs with respect to their arithmetic averages, their absolute variability -- in dollars -- is greater.

If one recalls that in perhaps 60 percent of all DRGs in South Carolina, the standard deviation exceeded $200 in 1981 and that in Florida the aggregate standard deviation was $417, the implications for a DRG-related physician reimbursement policy based on average allowed charges are that, in a very large percentage of cases, there could be significant financial incentives for physicians to reduce the provision of care. There would be an equal number of cases where the incentives might be to increase care, with the attendent increase in beneficiary coinsurance burdens.

At the hospital level, there is some chance that these reimbursement changes will wash out. Even in the relatively small state of South Carolina, there were an average of roughly 1200 Medicare discharges per hospital in 1981. For individual physicians, there is a much smaller chance that changes in reimbursement will net out. Most physician practices will face only a relative handful of Medicare hospitalizations. Thus additional work on the potential redistributive effects and the potential consequences of such redistributions will be very important. Finally, with respect to the beneficiaries, there will be little or no chance that changes in their liability could be averaged out over several hospitalizations. Nearly seventy percent of those Medicare beneficiaries hospitalized in any one calendar year will have only one hospitalization. The percentage of beneficiaries with more than two hospitalizations is very small.

VARIATIONS IN PHYSICIAN COORDINATION

One other perspective on the question of variations in physician charges involves the different physicians involved in providing care to hospitalized beneficiaries. Some of the variations in charges will be due to differences in the number of physician practices involved in any episode as well as differences in the charges of each physician practice. HCFA's intramural analysis has shown that the average number of practice identifiers associated with physician services provided in an inpatient setting for hospitalized beneficiaries ranges between 2 and 4. For this study, an alternative aggregate measure was computed: the ratio of the charges of the lead physician for a particular episode to the

total charges of all physicians services provided in the hospital for a particular beneficiary. The lead physician was identified as the practice responsible for either the purality of the hospital visits associated with discharges in Medical DRGs, or the practice with the largest share of allowed charges for Surgical DRGs. This measure may shed some light on the potential coordination problems that might be faced in trying to allocate any reimbursements paid on an integral DRG basis rather than on the existing piecemeal fee-for-service basis.

In both states, the lead physician was indeed responsible for most of the allowed charges during the stay. In South Carolina, the lead physician was responsible for 74 percent of all allowed charges while in Florida the lead practice was responsible for only 63 percent. As might be expected, Surgical DRGs were more concentrated in payments of allowed charges than Medical DRGs. In South Carolina, the lead physician proportions were 76 and 73 percent, respectively, while in Florida the statistics were 71 and 60 percent.

In South Carolina, the range in the lead physician percentage of allowed charges across the high volume DRGs went from 58 percent to 91 percent. In Florida, the minimum percentage was 49 percent and the maximum was 86 percent. In South Carolina, the lead physician percentage was greater than 70 percent in 48 of the 67 studied DRGs. In Florida, 49 of 67 were less than 65 percent.

Where these ratios are on the order of 85 to 90 percent, payment of a DRG rate to a single entity, e.g., the lead physician practice, may be somewhat plausible since relatively few additional dollars would have to be funneled through that channel. Where 60 percent or fewer of the payments would be retained by some initial entity with the rest being reallocated possibly through some negotiation process, the overhead of coordination may become considerable and the incentives with respect to cost effective medical practice may become diluted. Under a system where payments were made to some greater collection of physician practices, appropriate incentives might be expected to be even further attenuated.

Outliers

In the course of the project, Mandex computed means, standard deviations, and coefficients of variation for a wide variety of variables using both the South Carolina and Florida data. A substantial range was observed for the values of coefficients of variation. This was initially interpreted as implying substantial variability in some of those variables. However, visual inspection of plots of those variables revealed that apparent differences in variability virtually always were the result of the presence of observations which were statistical outliers. In fact, comparisons of the normalized cumulative distributions of a selection of DRGs which varied with respect to their CV values revealed substantial similarity between the bulk of the distributions across the DRGs for both Part B charges and Part A charges.

These visual inspections also revealed that, as expected, there is a strong association between Part A charges and Part B charges. Regression equations of the form:

Part B charges = constant + (multiplier) x (Part A charges) + error

were estimated for a selection of DRGs. The t-statistics on the estimated value of the multiplier term were always highly significant. (This finding of covariability implies that one cannot meaningfully compare coefficient of variation measures for Part A charges and Part B charges for individual DRGs.)

These plots also revealed that some observations which were "outliers" with respect to Part A charges were also outliers with respect to Part B charges. However, there were also observations which appeared to be outliers with respect to only Part B but not Part A, and others that were outliers only for A and not B. Further, because of the positive slope of the estimated regression lines, some observations with relatively high values for one set of charges and low values for the other appeared to be beyond the tolerance of confidence ellipses which included observations which had relatively high or low values for both sets of charges.

These tentative "lessons" surfaced a different question that must be resolved if a physician DRG payment system is to be implemented: how will physician charge outliers be identified? Retrospectively, one can identify an outlier from the set of history claims once a mean and a tolerance are specified. Prospectively one might be able to specify means and tolerances based on a previous year's values, but in a physician DRG payment system, there may not be a stream of claims that would be accumulated and that would eventually pass the outlier threshold. There may only be a single claim for all services rendered for a patient in a particular DRG. If such a system were implemented, there might be an incentive for physician practices to submit a bill at or above the threshold for any relatively expensive course of treatment that was approximately equal to the threshold. Alternatively, if Part B outliers were to be identified only when they met the Part A outlier criteria, some Part B outlier treatments would be under-reimbursed, some would be over-reimbursed, and hospitals and physicians might have the same financial incentive to continue to expend resources for patients whose course of treatment approached the outlier limits.

Whether these possible sets of incentives could vitiate the general prospective payment incentives for lower cost treatments would be an empirical question. The programmatic question of outlier identification, however, will need to be resolved.

Single versus Multiple Hospitalizations

At the outset of the study, the question arose whether the charges of beneficiaries with a single hospitalization differed from those of beneficiaries with multiple hospitalizations. It was conceivable that beneficiaries with multiple hospitalizations were sicker individuals. If that were to be verified, a separate reimbursement policy might be possible with respect to the second and any subsequent hospitalizations for an individual beneficiary. On the other hand, if there were no differences between the charges associated with single stays and those of multiple stay patients, the data needed for computation of potential MD DRG rates could be halved. Just under 50 percent of the

Medicare stays in any one calendar year derive from beneficiaries who are hospitalized only once during that year. Using the data of beneficiaries with only a single stay would save some effort, and might allow reliance on the HCFA inhouse Bill Summary Record data available for a sample of 5 percent of all beneficiaries.

Based on the estimated means and standard deviations from a selection of 67 high volume DRGs, t values were computed to compare by DRG the average physician allowed charges between beneficiaries with single hospitalizations and those with multiples. (Hospital stays included in the "multiple" category were all stays of any patient with more than one stay in 1981, not just those of patients with two or more stays in the same DRG.) Based on those values, there appears to be no evidence that patients with multiple stays are consistently associated with higher physician allowed charges. Half of the selected DRGs had higher physician charges for the single hospitalizations, half had higher values for the multiples.

As a result, there appears to be little justification for a separate set of reimbursement rates for episodes involving beneficiaries with more than one hospitalization. Since any apparent single hospitalization may be part of a multiple set of hospital stays spread out over more than one calendar year, it may be just as well that only one set of rates could suffice. However, although the singles and multiples are not consistently different, there are a sufficient number of significant differences in average allowed charges that one could not use only the singles or only the multiples to compute reliable relative rates for a physician DRG system.

SUMMARY

The "lessons" we have gleaned neither confirm the feasibility of a physician reimbursement system based on DRGs nor do they refute the possibility of such a system. The results of this study have identified a variety of policy decisions that will have to be made if a physician DRG system were to be implemented. The data can refine the issues but as yet, they do not dictate the

answers. Those in authority in either HHS or the Congress will have to make those decisions.

The unresolved computational issues are not particularly significant. DRG specific, single relative rates based on averages or other central tendency measures can be computed with some confidence to reflect billing data available to Medicare and its carriers and intermediaries. There is some question whether these data entirely reflect medical practice that might be observed among Medicare beneficiaries, but data filtering and trimming of outliers should produce a series of reasonable relative values by DRG.

The greater uncertainties in such a system would involve the potential magnitudes of changes in reimbursement and the resulting changes in physician behavior with respect to clinical practice. The clinical heterogeneity in some DRGs suggests that some types of treatments might be uniformly under-reimbursed given the existing patterns of practice. The variation observed in physician allowed charges suggests that some physicians and many beneficiaries would experience significant changes in Medicare financial coverage. The variation in the average proportion of allowed charges associated with the lead physician practice suggests that there may be significant coordination problems associated with trying to implement DRG payments to a single physician entity; single payments to a collective physician entity, at best, might leave incentives for cost efficiency severely attenuated.

Useful work which could be pursued at this point would examine and estimate the potential redistributive effects of an MD DRG system across all DRGs. Further, physician review of the clinical heterogeneity in many DRGs and the potential need for multiple physician reimbursement rates within DRGs would be essential. In addition, what is needed will be explicit identification and consideration of mechanisms which would both facilitate coordinated physician reimbursements and promote cost conscious clinical ordering behavior by physicians.